Title: A reliability model for the detection of electronic watermarks in digital images
Authors: Jean-Paul Linnartz, Ton Kalker, Geert Depovere, Rob Beuker
Conference: Fifth Symposium on Communications and Vehicular Technology in the Benelux 97
Date and location: October 97, Enschede, The Netherlands

# Form SF298 Citation Data

| Report Date ("DD MON YYYY") 01101997 | Report Type N/A | Dates Covered (from... to) ("DD MON YYYY") |
|---|---|---|

| | |
|---|---|
| **Title and Subtitle** A reliability model for the detection of electronic watermarks in digital images | **Contract or Grant Number** |
| | **Program Element Number** |
| **Authors** | **Project Number** |
| | **Task Number** |
| | **Work Unit Number** |
| **Performing Organization Name(s) and Address(es)** IATAC Information Assurance Technology Analysis Center 3190 Fairview Park Drive Falls Church VA 22042 | **Performing Organization Number(s)** |
| **Sponsoring/Monitoring Agency Name(s) and Address(es)** | **Monitoring Agency Acronym** |
| | **Monitoring Agency Report Number(s)** |
| **Distribution/Availability Statement** Approved for public release, distribution unlimited | |
| **Supplementary Notes** | |
| **Abstract** | |
| **Subject Terms** | |
| **Document Classification** unclassified | **Classification of SF298** unclassified |
| **Classification of Abstract** unclassified | **Limitation of Abstract** unlimited |
| **Number of Pages** 11 | |

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 074-0188*



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 10/1/97 | Report |

**4. TITLE AND SUBTITLE**
A reliability model for the detection of electronic watermarks in digital images

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**
Not provided

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Information Assurance
Technology Analysis Center
(IATAC)
3190 Fairview Park Drive
Falls Church, VA 22042

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Defense Technical
Information Center
DTIC-AI
8725 John J. Kingman Road,
Suite 944

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**

**12b. DISTRIBUTION CODE**
A

**13. ABSTRACT** ***(Maximum 200 Words)***

A watermark is a perceptually unobstructive mark embedded in an image, audio or video clip or other multimedia asset. A watermark can carry additional information, for instance about the source and copyright status of a document or its intended recipient, its rights and restrictions. We analyze the reliability of detecting such watermarks, modeling it as a detection problem where the original content acts as noise or interference. Probabilities of incorrect detections are expressed in terms of the watermark-to-image power ratio, showing a significant similarity in the problem of detecting watermarks and that of receiving weak spread-spectrum signals over a radio channel with strong interference. Theoretical results are verified by experiments.

**14. SUBJECT TERMS**
Digital Images

**15. NUMBER OF PAGES**

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

# A reliability model for the detection of electronic watermarks in digital images

Jean-Paul M. G. Linnartz* A.A.C. Kalker* G.F.G. Depovere*
R.A. Beuker*

## Abstract

A watermark is a perceptually unobstructive mark embedded in an image, audio or video clip or other multimedia asset. A watermark can carry additional information, for instance about the source and copyright status of a document or its intended recipient, its rights and restrictions. We analyse the reliability of detecting such watermarks, modeling it as a detection problem where the original content acts as noise or interference. Probabilities of incorrect detections are expressed in terms of the watermark-to-image power ratio, showing a significant similarity in the problem of detecting watermarks and that of receiving weak spread-spectrum signals over a radio channel with strong interference. Theoretical results are verified by experiments.

## I Background

Electronic watermarking is a new research area, combining aspects of digital signal processing, cryptography, statistical communication theory and human perception. It aims at embedding additional data into clear content (images, audio etc) in a way that is difficult to remove [1-13]. Principal applications of electronic watermarks are in copyright enforcement, automatic metering and monitoring of asset usage in multi-media applications, piracy tracing, and in providing additional information, such as image captions. New multi-media networks and services facilitate the distribution of content, but at the same time make copying and copyright piracy simple. Here we see a clear need to embed copyright data, such as the ownership or the identity of the authorized user in an indelible way. Typical requirements for watermarking and fingerprinting methods include:

1. Erasing the watermark should be technically difficult. Methods should be robust against attackers knowing the watermarking algorithm but not the *key*.
2. Replacing the watermark by another watermark should be a difficult task.
3. The watermarking scheme should be robust to transmission and storage imperfections (such as compression, noise addition, format conversion, bit errors) signal processing artefacts (noise reduction, filtering).
4. It should be robust against typical attacks, e.g. described in [13].
5. It should also be robust against colluding pirates who combine multiple versions of the same content that are stamped with different watermarks.
6. The watermark should be unobstructive, and not annoying to bona-fide users.

The organization of the paper is as follows. Section I provides an introduction to the problem of watermarking, its potential and limitations. Section II introduces our model of the image and the watermark. It extends the idea proposed in [1] to regard the original content as interference during the detection of a weak wanted signal (namely the watermark). Section III reviews several proposed methods for watermarking and verifies where the model applies. Section IV develops the theoretical performance of a correlator detector, as typically proposed in many recent papers. The model is verified with experiments in Section V. Section VI concludes this paper.

*Authors are with Philips Natuurkundig Lab (Philips Research), WY 8.1, Holstlaan 4, 5656 AA, Eindhoven, the Netherlands, tel: +31 (40) 2742302, fax: +31 (40) 2744660, e-mail: (linnartz, kalker, depovere, beuker)@natlab.research.philips.com

# II Formulation of Model

We approach the problem of watermark detection by assuming a stationary process $p$ as a model for our set of images. Given a watermark $w(n)$ we decide whether an image is watermarked or not by computing a decision variable $y$ and comparing $y$ to a threshold $y_{thr}$. We will derive expressions for the statistical properties of $y$ and the reliability of detection.

## II.1 Image Model

We address an image of size $N_1$ by $N_2$ pixels with a total of $N = N_1 N_2$ pixels. The intensity level of the pixel with coordinates $n = (n_1, n_2)$, $(0 \leq n_1 \leq N_1 - 1, 0 \leq n_2 \leq N_2 - 1)$ is denoted as $p(n)$. We denote $e_1 = (1, 0)$ and $e_2 = (0, 1)$, so $n = n_1 e_1 + n_2 e_2$. The set of all pixel coordinates is denoted as $A$, where

$$A = \{n : 0 \leq n_1 \leq N_1 - 1, 0 \leq n_2 \leq N_2 - 1\}.$$

In color pictures, $p(n)$ is a YUV or RGB vector, but for sake of simplicity we restrict our discussion to gray scale images, in which $p(n)$ takes on real or integer values in a certain interval. Whenever convenient we will represent $p(n)$ as a $z$-expression $p(z)$ defined by

$$\begin{aligned} p(z) &= \sum_{n \in A} p(n) z^{-n} \\ &= \sum_{n \in A} p(n) z_1^{-n_1} z_2^{-n_2} \end{aligned}$$

The recipient of an image $q(n)$ sees an altered (e.g. watermarked, filtered, quantized or otherwise manipulated) matrix that resulted from $p(n)$. A watermark detector has to operate on the a posteriori observation $q(n)$ while having knowledge on the a priori statistical behaviour of $p(n)$. We model the image as a sample (or a statistical *realization*) of a random matrix of $N_1$ by $N_2$ values. The $k$-th moment of the gray level of each pixel is denoted as $\mu_k = \mathrm{E}[\, p^k(n)\,]$. For our analysis, we assume spatial stationary, thus $\mu_k$ is considered not to depend on the location of the pixel. In particular, $\mu_1$ represents the average value or *DC-component* in the image and $\mu_2 = \mathrm{E}[\, p^2(n)\,]$ represents the average power in a pixel and $E_p = N\mu_2$ is the average total energy in an image. The variance is $\sigma^2 = \mathrm{E}[\,(p(n) - \mu_1)^2\,] = \mu_2 - \mu_1^2$. The intensity levels of pixels $n$ and $m = (m_1, m_2)$ are correlated, with

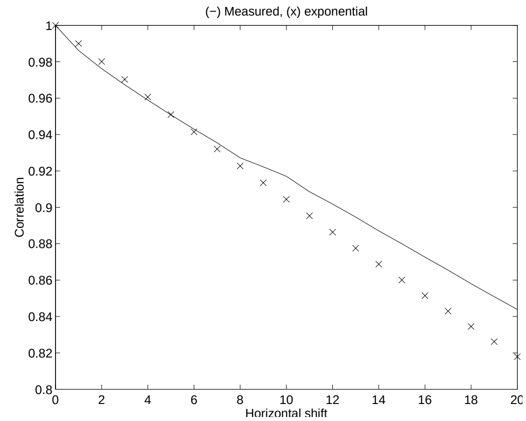


Figure 1: Normalized Correlation $(R_p(\Delta) - \mu_1^2)/\sigma^2$ versus horizontal shift $\Delta_1$ expressed in pixels. Solid line: measurement for "Teeny 1" image. Crosses: theoretical results for $\alpha = 0.99$.

$$\mathrm{E}[\, p(n)p(m)\,] = R_{p,p}(n - m).$$

The correlation only depends on the difference vector $\Delta = (\Delta_1, \Delta_2) = (n_1 - m_1, n_2 - m_2)$, as we assume that the image has homogeneous statistical properties (wide-sense spatial stationarity). In order to make calculations for our examples tractable, we simplify the image model assuming the first-order separable autocorrelation function (*acf*) [14]

$$R_{p,p}(\Delta) = \mu_1^2 + \sigma^2 \alpha^{|\Delta|}$$

where $|\Delta| = |\Delta_1| + |\Delta_2|$ and where $\alpha$ can be interpreted as a measure of the correlation between adjacent pixels in the image. Experiments reveal that typically $\alpha \approx 0.9 \ldots 0.99$. As illustrated in Figure 1, we found $\alpha \approx 0.99$ for the *Teeny 1* image. For this image, the sample mean is $\mu_1 = 103$ and the sample standard deviation is about $\sigma = 53$. We denote $\tilde{p}(n)$ as the non-DC components of the image, that is $\tilde{p}(n) = p(n) - \mu_1$, so $R_{\tilde{p}} = \sigma^2 \alpha^{|\Delta|}$. Some of the assumptions made here seem a crude approximation of the typical properties of images. From experiments such as those to be reported in Section V, it appeared that reliability estimates based on this crude model can be reasonably accurate for the purpose of this evaluation. These assumptions, however, exclude certain images, such as binary images or computer-generated images with a limited number of colors.

## II.2 Watermark Model

To detect a watermark in a suspect image, some proposed methods only use the suspect image and reference data on the watermark, while other methods also require the availability and use of the original image. We assume here that $p(n)$ is not available at the detector. The watermark is represented by $w(n)$ which takes on real values in all pixels $n \in A$. This watermark $w(n)$ is added to the original image. This results in the marked image $q(n) = p(n) + w(n)$. This model implicitly assumes that no spatial transformation of the image (resizing, cropping, rotation, etc.) is conducted. We aim at detecting whether a particular watermark is present or not, based on knowledge of $w(n)$.

In the special case that the watermark does not depend on $p(n)$, i.e., if $w(n)$ would be identical for any image $p(n)$, we may define signal-to-noise ratios (watermark energy-to-pixel variance), and we can apply theoretical frameworks from detecting communication signals transmitted over noisy channels. The correlation of watermark $w_1$ with $w_2$ is

$$R_{w_1,w_2}(\Delta) = \frac{1}{N} \sum_{n \in A} w_1(n) w_2(n + \Delta)$$

where we assume for simplicity that $n+\Delta$ wraps around when it formally falls outside of the set $A$. Note that the correlation $R_{w_1,w_2}$ is defined by a spatial summation whereas the correlation $R_{p,p}$ is defined by the statistics of the random process $p(n)$. If the image size is large enough ($N1 \gg 0, N2 \gg 0$) and if the process $p(n)$ is assumed to be ergodic we are allowed to approximate the statistical autocorrelation $R_{p,p}(\Delta)$ by a spatial autocorrelation

$$R_{p,p}(\Delta) \approx \frac{1}{N} \sum_{n \in A} p(n) p(n + \Delta).$$

The average energy in a watermark $w$ equals $E_w = R_{w,w}(0)$ and is denoted as $E_w$.

The DC content of the watermark $w$ is $W_0(w) = \frac{1}{N} \sum_{n \in A} w(n)$. A watermark $w$ is DC-free if $W_0(w) = 0$.

Some watermarks are generated by randomly generating a $+k$ or $-k$ pixel value for $w(n)$, independently for each pixel $n$. When this process is ergodic and the image size is large enough we may assume $R_{w,w}(\Delta_1) = R_{w,w}(\Delta_2)$ for $\Delta_1, \Delta_2 \neq 0$. With this assumption it follows that

$$R_{w,w}(\Delta) = \begin{cases} E_w & \text{if } \Delta = 0 \\ \frac{1}{N-1}(NW_0^2 - E_w) & \text{otherwise} \end{cases}. \tag{1}$$

By the law of large numbers $W_0^2$ decays with $\frac{1}{N}$. The non-zero terms in the autocorrelation function are therefore of the form , $/N$. This implies that for $N_1, N_2$ large enough the autocorrelation function $R_{w,w}$ will approximate a $\delta$-function. Watermarks with this property are referred to as *white* watermarks. A watermark is called *purely white* if its autocorrelation function is exactly equal to a $\delta$-function. We have seen that purely white watermarks cannot be DC-free.

As an other example, we will treat the case that the watermark has a low-pass spatial spectrum. This method has for instance been advocated by Cox et al. [9]. In such situation, a potential attacker can not easily remove the watermark by low-pass filtering. Moreover, JPEG compression typically removes or distorts high-frequency components. A low-pass watermark can be generated by spatially filtering a spatially white watermark $w(z)$. A first-order two dimensional spatial smoothing IIR filter $S_\beta(z) = \left[(1 - \beta z_1^{-1})(1 - \beta z_2^{-1})\right]^{-1}$ computes $\hat{w}(z) = S_\beta(z) w(z)$. Considering only the zeroth-order approximation of $R_{w,w}$ (i.e. discarding the terms with postive powers of $\frac{1}{N}$) we may put $R_{w,w}(z) = E_w$. Then $R_{\hat{w},\hat{w}}(z)$ is computed as

$$\begin{aligned} R_{\hat{w},\hat{w}}(z) &= S(z) S(z^{-1}) R_{w,w}(z) \qquad (2) \\ &= E_w \sum_n \frac{\beta^{|n|}}{(1 - \beta^2)^2} z^{-n}. \end{aligned}$$

It follows that $\hat{w}$ has a first order *acf* with correlation factor $\beta$. The average energy of $\hat{w}$ is given by

$$E_{\hat{w}} = \frac{E_w}{(1 - \beta^2)^2}.$$

# III Review of some watermark proposals

Next, we will review a few proposals for watermarking methods, and discuss how these meth-

ods relate to our framework. Many watermarking methods have been proposed in which the watermark is linear and independent of the image, similar to our assumptions. It turns out that methods by Bender et al. [3] [10], Pitas et al. [4], Cox et al. [9] all fit into the framework covered here. Moreover their proposed detection closely resembles the correlator concept to be covered in the next section.

Bender et al. [3] [10] describe a watermarking method called "Patchwork". This method takes $n_p (n_p \leq N/2)$ pairs of image points in a way known to the transmitter and receiver. The brightness of a pixel is increased by one and the brightness of the corresponding point is decreased by one. That is, using our notation of Section II, $w(n) = -1$ in $n_p$ points, it equals 1 in $n_p$ points, and is 0 in the remaining $N - 2n_p$ points. Thus, the watermark is DC-free ($W_0 = 0$) and the average energy in the watermark is $E_w = 2n_p/N$. The authors use the $n_p$ points to detect the watermark, but ignore all other $N - 2n_p$ points. However, because of the correlation in pixel values, these other $N - 2n_p$ can be exploited in the detector, as we will show later. In [14], a communication-theoretical evaluation is presented, which use simular assumptions as in our Section II, however ignoring for instance correlation between pixel values.

Pitas and Kaskalis [4] describe a similar method of partitioning the set of pixels into three subsets of size $n_p, n_p$ and $N - 2n_p$, respectively. As a special case, one can take $n_p = N/2$. The two subsets of equal size are used to embed and detect the watermark. The brightness of the pixels of one subset is altered by adding a positive integer factor $k$, which depends on the image $p(n)$ such that the watermark-to-image ratio is sufficiently large. This method can be captured in our model. We have $E_w = k^2/2$, $W_0 = k/2$.

Cox et al. [9] embed a sequence of real numbers of length $n_p$ in an $N_1$ by $N_2$ image by computing the $N_1$ by $N_2$ DCT and adding the sequence to the $n_p$ highest DCT coefficients, excluding the DC component. The method is sensitive to errors in determining the relative strength of DCT components, so the original is used for this purpose by both the sender and the recipient.

On the other hand, Zhao and Koch [8] [5] propose a method that does not lead itself well to modelling by our framework. Their watermark embeds a bitstream in the DCT domain. The image is divided up into 8x8 JPEG blocks. In predefined 8x8 blocks which are known only to the sender and the watermark detector, the DCT coefficients are modified to ensure a certain relative size. In this scheme $w(n)$ substantially depends on $p(n)$, and in a nonlinear way.

# IV Correlator detector

Correlator detectors are interesting to study, for several reasons. They are a mathematical generalization of the simple device in which watermarks with $w \in \{, 0, +1\}$ are detected by computing the normalized sum of all pixel values in which the watermark is negative, i.e., $s_- = \frac{1}{N} \sum_{n:w(n)=-1} q(n)$ and the normalized sum of all pixel values in which the watermark is positive, i.e., $s_+ = \frac{1}{N} \sum_{n:w(n)=1} q(n)$. Then, $y = s_+ - s_-$ is used as a decision variable, e.g. [3] [10]. Moreover, correlators are known to be the optimum detector for particular situations, namely the Linear Time-Invariant (LTI), frequency non-dispersive, Additive Gaussian Noise (AWGN) channel, when the receiver has full knowledge about the alphabet of waveforms used to transmit a message.

In a correlator detector, a decision variable $y$ is extracted from the received image $q(n)$ by correlating with a locally stored copy of the watermark $w(n)$. Therefore $y = R_{w,q}(0)$, with

$$R_{w,q}(\Delta) = \frac{1}{N} \sum_{n \in A} w(n) q(n + \Delta).$$

Figure 2 illustrates this correlation detector. The model covers all detectors in which the decision variable is a linear combination of pixel luminance values in the image. Hence, it is a generalization of many detectors proposed previously. It covers a broader class of watermarks then the binary ($w(n) \in \{-k, k\}$) or ternary ($w(n) \in \{-k, 0, k\}$) watermarks. It particularly includes methods in which watermark data is added to DCT coefficients. For our analysis, we separate $y$ into a deterministic contribution $y_w$ from the watermark,

$$\begin{aligned} y_w &= \frac{1}{N} \sum_{n \in A} w(n) w(n) \\ &= R_{w,w}(0) \\ &= E_w \end{aligned}$$

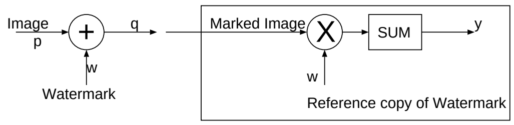


Figure 2: Watermark Embedder and Correlation Detector

plus filtered noise from the image $y_p$

$$\begin{aligned} y_p &= \frac{1}{N}\sum_{n\in A} w(n)p(n) \\ &= R_{w,p}(0). \end{aligned}$$

Regarding $y_p$, the mean value is found as the product of the DC component in the watermark and the image, with

$$\begin{aligned} \mathrm{E}[\,y_p\,] &= \frac{1}{N}\mathrm{E}[\sum_{n\in A} w(n)p(n)\,] \\ &= \frac{\mathrm{E}[\,p(n)\,]}{N}\sum_{n\in A} w(n) \\ &= \mu_1 W_0. \end{aligned}$$

This result appears to be irrespective of the correlation in pixels. To find the second moment, we compute

$$\begin{aligned} \mathrm{E}[\,y_p^2\,] &= \mathrm{E}[\left(\frac{1}{N}\sum_{n\in A} w(n)p(n)\right)^2] \qquad (3) \\ &= \frac{1}{N^2}\mathrm{E}[\sum_{n,m\in A} w(n)p(n)w(m)p(m)\,] \\ &= \frac{1}{N^2}\mathrm{E}[\sum_{\substack{n\in A\\ n+\Delta\in A}} q(n,\Delta)r(n,\Delta)\,] \\ &= \mu_1^2 W_0^2 + \\ &\quad \frac{1}{N^2}\left(\mathrm{E}[\sum_{\substack{n\in A\\ n+\Delta\in A}} q(n,\Delta)\tilde{r}(n,\Delta)\,]\right) \end{aligned}$$

where we have written $\tilde{p}(n) = p(n) - \mu_1$, $q(n,\Delta) = w(n)w(n+\Delta)$, $r(n,\Delta) = p(n)p(n+\Delta)$ and $\tilde{r}(n) = \tilde{p}(n)\tilde{p}(n)$. Note that for correlated pixels ($\alpha > 0$) and spectrally non-white watermarks, one has to account for non-zero cross terms with $\Delta \neq 0$. For the remainder of this paper we will now assume that the image is sufficiently large and that the autocorrelation function $R_{\tilde{p},\tilde{p}}(\Delta)$ decays sufficiently fast. This allows us to drop the restriction $n + \Delta \in A$ in the summation of Equation 3. Continuing the computations from Equation 3 and assuming a stationary and ergodic image source we find

$$\begin{aligned} \mathrm{E}[\,y_p^2\,] - \mu_1^2 W_0^2 &= \qquad (4) \\ &= \frac{1}{N^2}\left(\sum_{\Delta}\sum_{n\in A} q(n,\Delta)\mathrm{E}[\,\tilde{r}(n,\Delta)\,]\right) \\ &= \frac{1}{N^2}\left(\sum_{\Delta}\sum_{n\in A} q(n,\Delta)R_{\tilde{p},\tilde{p}}(\Delta)\right) \\ &= \frac{1}{N}\left(\sum_{\Delta} R_{w,w}(\Delta)R_{\tilde{p},\tilde{p}}(\Delta)\right). \end{aligned}$$

It follows that the variance $\sigma_d$ of the decision variable $y$ due to image noise is given by the expression

$$\sigma_d^2 = \frac{1}{N}\sum_{\Delta} R_{w,w}(\Delta)R_{\tilde{p},\tilde{p}}(\Delta). \qquad (5)$$

In order to have a reliable detection of a watermark it is required that the ratio $E_w/\sigma_d$ is as large as possible. We therefore define the reliability of detection $\rho_d$ as the ratio $E_w/\sigma_d$. The consequences of Equation 5 with respect to $\rho_d$ will be elaborated in the next sections.

## IV.1 Example 1: White watermarks

The white watermark reasonably models most of the early proposals for increasing and decreasing the pixel luminance according to a pseudo random process. For an absolutely DC-free watermark, $W_0(w) = 0$, the decision variable $y$ will have zero mean.

For a white binary watermark with embedding depth $k$, the variance $\sigma_d^2$ is computed as

$$\begin{aligned} \sigma_d^2 &= \frac{E_w\sigma^2}{N} + \frac{\sigma^2,}{N(N-1)}\sum_{\Delta\neq 0}\alpha^{|\Delta|} \\ &= \frac{E_w\sigma^2}{N} + \frac{\sigma^2,}{N(N-1)}\left[\left(\frac{1+\alpha}{1-\alpha}\right)^2 - 1\right] \end{aligned}$$

where , is some constant dependent on the watermark generation process (see Section II.2). We see that the effect of pixel correlations is significant only if $\alpha$ is very close to unity (little luminance changes) in a small-size image. If the image is large enough, that is, if $N \gg \left[\frac{1+\alpha}{1-\alpha}\right]^2$,

we may approximate

$$\sigma_d^2 \approx \frac{E_w \sigma^2}{N}$$

In practical situations, this appears a reasonable approximation. It follows that $\rho_d$ is given by

$$\rho_d = \frac{\sqrt{E_w N}}{\sigma}. \qquad (6)$$

Note that Equation 6 implies that the reliability only depends on the total energy of the watermark and not on how this energy is distributed.

The value of $\rho_d$ in Equation 6 will be used in Section IV.3 to obtain an expression for the probability of an incorrect detection.

## IV.2 Example 2: Low-pass watermark

We now address a watermark with independently randomly chosen pixel values, but which is then filtered by a two-dimensional first-order filter $S_\beta(z)$. Considering only the zero order terms and using the results from Section II.2 (Equation 2) we compute

$$\begin{aligned} \sigma_d^2 &= \frac{E_w \sigma^2}{N} \sum_{\Delta} (\alpha\beta)^{|\Delta|} \\ &= \frac{E_w \sigma^2}{N} \left[ \frac{1+\alpha\beta}{1-\alpha\beta} \right]^2 \end{aligned}$$

This reduces to the result of Section IV.1 for $\beta \to 0$, i.e. for a white watermark. The reliability is given by

$$\rho_d = \frac{\sqrt{E_w N}}{\sigma} \frac{1-\alpha\beta}{1+\alpha\beta}. \qquad (7)$$

Comparing this with Equation 6 we find that the use of low-pass watermark leads to a loss of reliability for $\alpha \neq 0$.

## IV.3 Error Rate

Because of the Central Limit Theorem, $y_p$ has a Gaussian distribution if $N_1, N_2$ are sufficiently large and if the contributions in the sums are sufficiently independent. The Gaussian behaviour will be verified in Section V. If we apply a threshold $y_{thr}$ to decide that the watermark is present if $y - \mathrm{E}[y_p] > y_{thr}$, the probability of a *missed detection* (the watermark is present in $q(n)$, but the detector thinks it is not) is

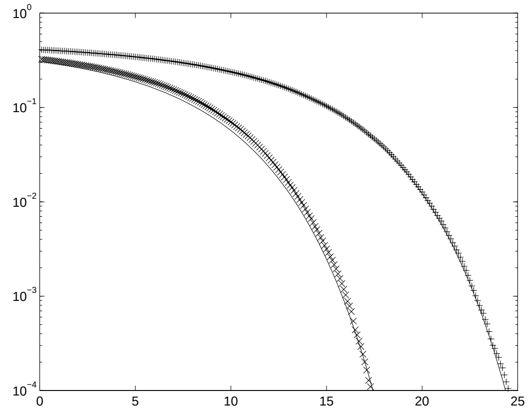


Figure 3: Watermark detection error rates $P_{fa}$ and $P_{md}$ versus signal-to-noise ratio $E_w/\sigma^2$ for correlation detector. Experiments on *Lenna* ($\times$). Solid line: corresponding theoretical curve.

$$P_{md} = \frac{1}{2} \operatorname{erfc}\left( \frac{E_w - y_{thr}}{\sqrt{2}\sigma_d} \right).$$

On the other hand, given that no watermark is embedded, a *false alarm* occurs with probability

$$P_{fa} = \frac{1}{2} \operatorname{erfc}\left( \frac{y_{thr}}{\sqrt{2}\sigma_d} \right).$$

Putting $y_{thr} = E_w/2$ provides

$$\begin{aligned} P_{md} &= P_{fa} \qquad (8) \\ &= \frac{1}{2} \operatorname{erfc}\left( \frac{E_w}{2\sqrt{2}\sigma_d} \right) \\ &= \frac{1}{2} \operatorname{erfc}\left( \frac{\rho_d}{2\sqrt{2}} \right) \end{aligned}$$

For the low-pass type of watermark of Section IV.2 the error rate goes into

$$\begin{aligned} P_{md} &= P_{fa} \\ &= \frac{1}{2} \operatorname{erfc}\left( \sqrt{\frac{E_w N}{8\sigma^2}} \frac{1-\alpha\beta}{1+\alpha\beta} \right) \end{aligned}$$

This result clearly shows that if the watermark is confined to low-pass components of the image, this significantly affects the reliability of detection. In this case the random +/- terms in $y_p$, which are due to multiplying the image $p$ with the locally stored copy of the watermark $\hat{w}$, do

not cancel as rapidly as these would vanish for a white watermark. If the watermark contains relatively strong low-frequency components (large $\beta$), the variance of $y_p$ is stronger and the error rate is larger. If the watermark contains relatively strong high-frequency components $\beta \approx 0$, the variance is weaker, so the watermark sees less interference from the image itself. However, such a high-frequency watermark is more vulnerable to erasure by image processing, such as low-pass filtering (smoothing).

# V Computational and Experimental Results

In our experiments, we approximated white watermarks through pseudo-random sequences. An appropriate choice appeared to be binary watermarks, $w(n) \in \{-k, k\}$ with $\beta = 0$ generated by a 2-dimensional LFSR maximal length sequence [16] [17] of length $2^{14} - 1 = 127 * 129$. Such sequences have a negligibly small DC component $\sum_n w(n) = -1$ and a correlation function that has the appropriate $\delta$-function shape. Repetition of the 127 by 129 basic pattern leads to a periodic correlation function, but maintains virtually zero correlation outside the peaks.

Figure 3 compares the above theoretical results with measurements of the *Lenna* and *Teeny* image. This Figure shows a good agreement between theory and experiment. In order to get statistical results we simulated sources of images by taking shifted and noisy versions of one of these prototype images. We computed the components of the decision variable and estimated which signal-to-noise ratio would be needed to achieve reliable detection.

# VI Conclusions

In this paper, we proposed a mathematical framework to model electronic watermarks embedded in digital images. The model regards the process of embedding and detecting a watermark to be similar to that of a communication channel. It treats the original contents (the image itself) as interference noise.

We observe that many detectors proposed for watermarks are of the correlator type, though often with minor modifications. Several essential differences appear with the case of transmission over a linear time-invariant channel with AWGN. Our model predicts reliability performance (missed detection and false alarms). In some special cases, particularly that of a white watermark, the signal-to-noise ratio (watermark-to-content-energy) appears the only factor to influence the reliability of detection. This leads to expressions for error probabilities similar to those experienced in radio communication. However, the spectral content of the watermark appears another critical parameter. If the watermark is non-white, the spectral properties of the images have a significant influence.

# References


[1] B.M. Macq and J.J. Quisquater, "Cryptology for digital tv broadcasting" Proc. of the IEEE, Vol. 83 No. 6, 1993, pp. 944-957

[2] R.G. van Schyndel, A.Z. Tirkel, C.F. Osborne: "A Digital Watermark", Int. IEEE Conf on Image Processing, Vol.2., 13-16 Nov. 1994, IEEE Comput. Soc. Press, Los Alamitos, CA, USA, pp. 86-90

[3] W. Bender, D. Gruhl, N. Morimoto, "Techniques for Data Hiding", Proceedings of the SPIE, 2420:40, San Jose CA, USA, February 1995

[4] I. Pitas, T. Kaskalis : "Signature Casting on Digital Images", Proceedings IEEE Workshop on Nonlinear Signal and Image Processing, Neos Marmaras, June 1995

[5] E. Koch, J. Zhao : "Towards Robust and Hidden Image Copyright Labeling". Proceedings IEEE Workshop on Nonlinear Signal and Image Processing, Neos Marmaras, June, 1995

[6] Caronni G.: "Assuring Ownership Rights for Digital Images", Proceedings of Reliable IT Systems, VIS '95, Vieweg Publishing Company, Germany, 1995

[7] F.M. Boland, J.J.K. O Ruanaidh, C. Dautzenberg, "Watermarking Digital images for Copyright Protection", Proceedings of the 5th IEE International Conference on Image Processing and its Applications, no. 410, Edinburgh, July, 1995, pp. 326-330.

[8] J. Zhao, E. Koch : "Embedding Robust Labels into Images for Copyright Protection", Proceedings of the International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Vienna, Austria, August 1995

[9] I.J. Cox, J. Kilian, T. Leighton, T. Shamoon "Secure Spread Spectrum Watermarking for Multimedia", NEC Research Institute, Technical Report 95 - 10

[10] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding", IBM Systems Journal, Vol. 35. No. 3/4 1996

[11] I. Cox, J. Kilian, T. Leighton and T. Shamoon, "A secure, robust watermark for multimedia", in Proc. Workshop on Information Hiding, Univ. of Cambridge, U.K., May 30 - June 1, 1996, pp. 175-190

[12] J.R. Smith, B. O. Comiskey, "Modulation and Information Hiding in Images", in Proc. Workshop on Information Hiding, Univ. of Cambridge, U.K., May 30 - June 1, 1996, pp. 191 - 201

[13] I.J. Cox and J.P.M.G. Linnartz, "Public watermarks and resistance to tampering", accepted for presentation at Int. Conf. on Image Processing (ICIP) 1997.

[14] N.S. Jayant and P. Noll., "Digital Coding of waveforms" Prentice Hall, 1984.

[15] Ch. W. Therrien, "Discrete Random Signals and Statistical Signal Processing" Prentice Hall, 1992.

[16] F.J. McWilliams and N.J.A. Sloane, "Pseude-Random Sequences and arrays", Proc . of IEEE, Vol. 64, No.12, Dec. 1976, pp. 1715-1729

[17] D. Lin and M. Liu, "Structure and Properies of Linear Recurring m-arrays", IEEE Tr. on Inf. Th., Vol. IT-39, No. 5, Sep. 1993, pp. 1758-1762

**Biography** Ton Kalker was born in The Netherlands in 1956. He received his M.S. degree in mathematics in 1979 from the University of Leiden, The Netherlands. From 1979 until 1983, while he was a Ph.D. candidate, he worked as a Research Assistant at the University of Leiden. From 1983 until December 1985 he worked as a lecturer at the Computer Science Department of the Technical University of Delft. In January 1986 he received his Ph.D. degree in Mathematics. In December 1985 he joined the Philips Research Laboratories Eindhoven. Until January 1990 he worked in the field of Computer Aided Design. He specialized in (semi-) automatic tools for system verification. Currently he is a member of the Digital Signal Processing group of Philips Research. His research interests include wavelets, multirate signal processing, motion estimation, psycho physics, digital watermarking and digital video compression.